THE INSTITUTE OF STATISTICS

THE CONSOLIDATED UNIVERSITY OF NORTH CAROLINA

A New Result on Interpreting Lagrange Multipliers as Dual Variables†

by

F. J. Gould* *and* Stephen Howe*

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 738

January, 1971

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

18

† *This research was partially supported by a grant from the Office of Naval Research contract number N00014-67-A-0321-0003 (NR047-095).*

* *Curriculum in Operations Research and Systems Analysis, University of North Carolina at Chapel Hill, 27514.*

A NEW RESULT ON INTERPRETING LAGRANGE MULTIPLIERS AS DUAL VARIABLES†

by

F. J. Gould* *and* Stephen Howe*

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 738

January, 1971

A New Result on Interpreting Lagrange Multipliers as Dual Variables†

F. J. Gould* *and* Stephen Howe*

I. Introduction

It is well known that in linear programming the knowledge of optimal dual variables can be used to simplify the finding of an optimal solution to the primal problem. In this paper, the analogous notion is considered for nonlinear and particularly nonconcave programming problems. A new theoretic result on Lagrange multipliers is presented which enables one to make a statement for nonlinear problems similar to the above assertion for linear programs.

Consider the problem

$$\text{(P):} \qquad \text{maximize } f(x), \text{ subject to } \quad g(x) \le 0,$$

where $f: R^n \to R$, $g: R^n \to R^m$. Let S_0 represent the constraint set (feasible points) for this problem. That is

$$S_0 = \{x \in R^n: g(x) \le 0\}.$$

Assume that all functions in (P) are twice continuously differentiable and that (P) has at least a local solution, say x^*. That is, x^* is in S_0 and there exists a neighborhood of x^*, say $N(x^*)$, such that $f(x^*) \ge f(x)$ whenever x is in $N(x^*) \cap S_0$. Assume that a constraint qualification holds at x^* [2]. Then the following Kuhn-Tucker optimality conditions are valid:

---

† *This research was partially supported by a grant from the Office of Naval Research contract number N00014-67-A-0321-0003 (NR047-095).*

* *Curriculum in Operations Research and Systems Analysis, University of North Carolina at Chapel Hill, 27514.*

There exists a nonnegative vector λ^* in R^m such that

$$C_1: \quad \nabla f(x^*) = \sum_{i=1}^{m} \lambda^*_i \nabla g_i(x^*)$$

$$C_2: \quad \langle \lambda^*, g(x^*) \rangle = 0 \quad \text{(complementary slackness).}$$

In this context the following question may be posed. Suppose that the above λ^* vector is known, but that x^* is unknown. How might one then use the knowledge of λ^* to determine either x^* or possibly some other local solution to (P)?

If (P) is a concave program ($f(x)$ concave, $g_j(x)$ convex, $j=1,\ldots,m$) the answers to this question are known. In particular, in this case, the Kuhn-Tucker conditions are sufficient as well as necessary for optimality (i.e., if the pair (x^*,λ^*) satisfies C_1 and C_2 then x^* is a solution to (P). Also any x^* satisfying C_1 also (globally) maximizes the Lagrangian function $L(x,\lambda^*) = f(x)-\sum_{i=1}^{m} \lambda^*_i g_i(x)$, since for concave programs the Lagrangian is a concave function. If the Lagrangian is strictly concave, then x^* will be its unique global maximizer. If the Lagrangian is not strictly concave then there may be a set of vectors, including x^*, each member of which maximizes the Lagrangian (thus satisfying C_1). In this case, any feasible member of the set, which also satisfies C_2, is a solution to (P). Thus, whenever (P) is a concave program, a knowledge of λ^* permits the statement of a conceptual procedure which is necessary and sufficient for determining solutions to (P) in the sense that it is guaranteed to be failsafe. This procedure is the following. Find among the set of Lagrangian multipliers one that also satisfies feasibility and complementary slackness. The procedure is failsafe in that x^* is such a quantity.

If (P) is not concave, then there are at least two serious difficulties:

(1) There is no convenient way of determining the vectors $\hat{x}$ which

satisfy the system

$$D_1: \nabla f(x) = \sum_{i=1}^{m} \lambda^*_i \nabla g_i(x)$$

$$D_2: g(x) \leq 0$$

$$D_3: \langle \lambda^*, g(x) \rangle = 0.$$

In particular, in the nonconcave case, the global (local) solutions to (P) will not generally be global (local) maximizers of $f(x)-\sum_{i=1}^{m}\lambda^*_i g_i(x)$.

(2) The Kuhn-Tucker conditions in the nonconcave case are not generally sufficient. Hence, even if all solutions to the above system can be found, there is not a known way of selecting the solutions to (P) from the set of vectors $\hat{x}$ which satisfy D_1, D_2 and D_3. (Second order information could be used but there are no known necessary and sufficient second order conditions. Hence, conceptually, second order information does not solve the problem.)

Thus, in the case of nonconcave programs, (P), the usual Lagrangian function appears to be of limited use as a vehicle for answering the question: Can a knowledge of λ^* be theoretically related to a procedure for solving (P)?

In this paper, we consider the extended Lagrangian function [3]

$$(1.1) \qquad P(x,k,\beta) = f(x) - \sum_{i=1}^{m} \beta_i[exp(kg_i(x))-1],$$

where $x \in R^n$, k is a nonnegative scalar, and $\beta \in R^m$. Recall that x^* is some unknown local solution to (P), λ^* is a known nonnegative dual variable, and the pair (x^*,λ^*) satisfies the Kuhn-Tucker conditions D_1, D_2 and D_3. We now wish to assume that

(i): x^* satisfies the second order sufficiency conditions [4]

(ii): $\lambda^*_i > 0$ for every i such that $g_i(x^*) = 0$.

Under these conditions, it will be shown that there exists a positive scalar K such that if $k^* \geq K$, and if $\beta^*_i = \lambda^*_i/k^*$, $i = 1,\ldots,m$, then x^* is a local maximizer of $P(x,k^*,\beta^*)$, and any local maximizer of $P(x,k^*,\beta^*)$ also

satisfying D_2 and D_3 is a local solution to (P). Thus, from a knowledge of λ^*, we are able in theory to determine a local solution to (P) by the following failsafe procedure: conduct a search among the local maximizers of $P(x,k^*,\beta^*)$ for one which satisfies D_2 and D_3. The procedure is failsafe in that x^* is such an element. Thus, in nonconcave programs, although the usual Lagrangian function has restricted theoretic application, it is seen that the Lagrange multipliers, λ^*, continue to play a key theoretic role as dual variables. The use of an extended Lagrangian function allows one to extend the theory of concave programs in a parallel way.

In another, but related, context it will be shown that the function $P(x,k,\beta)$ actually has a local saddle point at x^*, k^*, β^*, and consequently the results herein have an interpretation in the framework of duality notions for nonlinear programs.

The present work has been motivated by a desire to improve the computational theory for solving nonconcave problems. Although immediate computational implications of the above described results are not clear, it is hoped that this work will lead to a better understanding of nonconcave programs and thereby stimulate other efforts to obtain computational advances.

In Section 2, the main result is stated and proved. Following a discussion in Section 3, the duality interpretation is demonstrated and discussed in Section 4. An example and application are presented in Sections 5 and 6, respectively.

II. Results

Let the problem (P) and x^* be defined as above. Suppose f and g are twice continuously differentiable, and that the weak constraint qualification

holds at x^*. Let I denote the index set of active constraints. That is,

$$I = \left\{i \in \{1,2,\ldots,m\}: g_i(x^*) = 0\right\}.$$

Let λ^* be a nonnegative vector such that (x^*,λ^*) satisfy D_1, D_2 and D_3 (i.e., λ^* is an optimal dual variable, or Kuhn-Tucker multiplier). Now define the extended Lagrangian $P(x,k^*,\beta^*)$, as in (1.1), by setting $k^* > 0$ and $\beta^*_i = \lambda^*_i/k^*$, $i = 1,2,\ldots,m$.

LEMMA 1. (x^*,λ^*) satisfy D_1,D_2,D_3 if and only if (x^*,k^*,β^*) satisfy

$$G_1:\quad 0 = \nabla P(x^*,k^*,\beta^*) = \nabla f(x^*) - \sum_{i=1}^{m} \beta^*_i \, exp(k^* g_i(x^*))k^*\nabla g_i(x^*)$$

$$G_2:\quad g_i(x^*) \leq 0, \quad i = 1,2,\ldots,m$$

$$G_3:\quad \beta^*_i = \beta^*_i \, exp(k^* g_i(x^*)), \quad i = 1,2,\ldots,m.$$

PROOF: For $i \notin I$, G_3 holds iff $\beta^*_i = 0$, iff $\lambda^*_i = 0$, iff D_3 holds. For $i \in I$, $g_i(x^*) = 0$ so that both G_3 and D_3 hold. Also, x^* is assumed to be a local solution to (P) and hence D_2 and G_2 (feasibility) are trivially satisfied. Thus the system G_1, G_2, G_3 is equivalent to

$$\text{(I)}\quad 0 = \nabla f(x^*) - \sum_{i\in I} k^* \beta^*_i \nabla g_i(x^*).$$

Similarly, the system D_1, D_2, D_3 is equivalent to

$$\text{(II)}\quad 0 = \nabla f(x^*) - \sum_{i\in I} \lambda^*_i \nabla g_i(x^*).$$

Finally, it follows from the assumption $\beta^*_i = \lambda^*_i/k^*$, each i, that (I) and (II) are the same. □

LEMMA 2. If a constraint qualification holds at x^* then for any positive k^* there exists a nonnegative vector β^* such that (x^*,k^*,β^*) satisfy G_1, G_2, G_3.

PROOF: Take $\beta^* = \lambda^*/k^*$. □

Now consider the Hessian of the extended Lagrangian at x^*:

$$(2.1)\qquad \nabla^2 P(x^*,k^*,\beta^*) = \nabla^2 f(x^*) - \sum_{i\in I} k^*\beta^*_i \nabla^2 g_i(x^*) - \sum_{i\in I} (k^*)^2 \beta^*_i \nabla g_i(x^*)\nabla^T g_i(x^*)$$

$$= \nabla^2 L(x^*,\lambda^*) - k^* \sum_{i\in I} \lambda^*_i \nabla g_i(x^*)\nabla^T g_i(x^*)$$

where $L(x,\lambda^*)$ is the usual Lagrangian function discussed in Section 1. Note that if $\nabla^2 L(x^*,\lambda^*)$ is negative definite, then $\nabla^2 P(x^*,k^*,\beta^*)$ is also, because in (2.1) each term of the sum is a dyadic matrix. Even when $\nabla^2 L(x^*,\lambda^*)$ is not negative definitive, however, the dyadic terms will under certain conditions cause $\nabla^2 P(x^*,k^*,\beta^*)$ to be negative definite. Let $B = \{i\in I: \lambda^*_i>0\}$, let S be the subspace spanned by $\{\nabla g_i(x^*): i\in B\}$, and denote the orthogonal complement of S by $S^\perp$.

THEOREM 1. Suppose that, in addition to D_1, D_2 and D_3, (x^*,λ^*) also satisfy D_4: for each nonzero $z\in S^\perp$, $z^T\nabla^2 L(x^*,\lambda^*)z < 0$. Then there is a positive scalar K such that for any $k^* > K$ and $\beta^*_i = \lambda^*_i/k^*$, $i = 1,2,\ldots,m$ the Hessian $\nabla^2 P(x^*,k^*,\beta^*)$ is negative definite. Consequently, x^* is an isolated local maximizer of $P(x,k^*,\beta^*)$.

REMARK: If the nondegeneracy assumption $\lambda^*_i > 0$, $i\in I$ is made, then assuming that (x^*,λ^*) satisfy D_1, D_2, D_3 and D_4 is equivalent to assuming that (x^*,λ^*) satisfy the second order sufficiency conditions discussed by Fiacco and McCormick in [4].

PROOF OF THEOREM 1. (i) If $S = \{0\}$ then $S^\perp = R^n$, in which case D_4 implies that $\nabla^2 L(x^*,\lambda^*)$, and hence $\nabla^2 P(x^*,k^*,\beta^*)$ for any positive k^*, is negative definite. In this case, the result holds for any $K > 0$.

If $S \neq \{0\}$, consider the continuous function $h: S \to R$ defined by

$$h(y) = y^T\left[\sum_{i\in B} \lambda^*_i \nabla g_i(x^*)\nabla^T g_i(x^*)\right]y = \sum_{i\in B} \lambda^*_i (y^T\nabla g_i(x^*))^2.$$

For any $y\in S$ such that $y \neq 0$, $h(y) > 0$, hence h takes a minimum $m_1 > 0$ on the compact set $U = \{y\in S: \|y\| = 1\}$. Now denote $M = \|\nabla^2 L(x^*,\lambda^*)\|$, the

sup norm of the matrix $\nabla^2 L(x^*,\lambda^*)$.

ii) If $S = R^n$, then for every $v \in S$ such that $v \neq 0$,

$$\begin{aligned} v^T \nabla^2 P(x^*,k^*,\beta^*) v &= v^T \nabla^2 L(x^*,\lambda^*) v - k^* h(v) \\ &\leq M\|v\|^2 - k^* m_1 \|v\|^2. \end{aligned}$$

In this case, the result holds for $K = M/m_1$.

iii) Finally, suppose $\{0\} \neq S \subsetneq R^n$. Then $S^\perp \neq \{0\}$. For any $z \in S^\perp \ni z \neq 0$, $z^T \nabla^2 L(x^*,\lambda^*) z < 0$, hence $z^T \nabla^2 L(x^*,\lambda^*) z$ has a maximum $-m_2 < 0$ on the compact set $V = \{z \in S^\perp : \|z\| = 1\}$. For any $v \in R^n$ $\exists y \in S$, $z \in S^\perp \ni v = y+z$. Then

$$\begin{aligned} v^T \nabla^2 P(x^*,k^*,\beta^*) v &= (y+z)^T \nabla^2 L(x^*,\lambda^*)(y+z) - \\ &\qquad (y+z)^T \left[k^* \sum_{i \in B} \lambda^*_i \nabla g_i(x^*) \nabla^T g_i(x^*) \right] (y+z) \\ &= z^T \nabla^2 L(x^*,\lambda^*) z + 2y^T \nabla^2 L(x^*,\lambda^*) z + y^T \nabla^2 L(x^*,\lambda^*) y - k^* h(y) \\ &\leq -m_2 \|z\|^2 + 2M\|y\|\,\|z\| + M\|y\|^2 - k^* m_1 \|y\|^2 \\ &= -k^* m_1 \|y\|^2 - m_2 (\|z\| - \frac{M}{m_2} \|y\|)^2 + \frac{M^2}{m_2} \|y\|^2 + M\|y\|^2 \\ &= \|y\|^2 (-k^* m_1 + \frac{M^2}{m_2} + M) - m_2 (\|z\| - \frac{M}{m_2} \|y\|)^2. \end{aligned}$$

Hence, if $k^* > K \equiv \frac{M}{m_1}(1 + \frac{M}{m_2})$, then $v^T \nabla^2 P(x^*,k^*,\beta^*) v < 0$ for any nonzero v.

□

III. Discussion

For the program P as defined, with local maximum at x^* and nonnegative vector λ^* such that (x^*,λ^*) satisfy the Kuhn-Tucker conditions, suppose λ^* is known. In the concave case, with no further assumptions, this information can be used to find a global solution to P. With the above results, we can now make a similar, but somewhat weaker, statement in the nonconcave case. Under the

additional assumption that D_4 holds, there is a value $K > 0$ such that if we set $k^* > K$ and $\beta^*_i = \lambda^*_i/k^*$ for $i = 1,\ldots,m$ then (x^*,k^*,β^*) will satisfy G_1, G_2 and C_3, and the Hessian of the extended Lagrangian $P(x,k^*,\beta^*)$ will be negative definite at x^*. This means that $P(x,k^*,\beta^*)$ will have a local unconstrained maximum at x^*. Furthermore, we can show that any local (global) maximizer of $P(x,k^*,\beta^*)$ which satisfies feasibility and complementary slackness will be a local (global) solution to P. For example, suppose $\hat{x}$ is a global maximizer of $P(x,k^*,\beta^*)$ and suppose $\hat{x}\epsilon S_0$ and that $(\hat{x},\lambda^*)$ satisfy the complementary slackness relations D_3. Then $\forall\, x\epsilon R^n$

$$f(\hat{x}) - \sum_{i=1}^{m} \beta^*_i[exp(k^*g_i(\hat{x}))-1] \geq f(x) - \sum_{i=1}^{m} \beta^*_i[exp(k^*g_i(x))-1],$$

which implies

$$f(\hat{x}) \geq f(x) + \sum_{i=1}^{m} \beta^*_i[exp(k^*g_i(\hat{x}))-exp(k^*g_i(x))]$$

$$= f(x) + \sum_{i=1}^{m} \beta^*_i[1-exp(k^*g_i(x))],$$

which implies $f(\hat{x}) \geq f(x)$ $\forall$ x in S_0. Hence $\hat{x}$ is a global solution to (P). The argument for local solutions is analogous.

Thus, given λ^* under the stated conditions, we formulate the extended Lagrangian in the form (1.1) and then search among its local maximizers for points $\hat{x}$ which also satisfy feasibility and complementary slackness. There will be at least one such point $\hat{x}$, namely x^*, and each such $\hat{x}$ will be a local solution to (P). Note that this procedure will not directly yield a global solution to (P), unlike the concave case. If the points $\hat{x}$ include a global solution to (P) (for example, if x^* is a global solution), it will be recognizable only to the extent that if x_0 is a global solution to (P) and if $\tilde{x}$ is any other local maximizer of $P(x,k^*,\beta^*)$ satisfying feasibility and complementary slackness, then $P(\tilde{x},k^*,\beta^*) \leq P(x_0,k^*,\beta^*)$.

IV. Duality

Under the assumptions in Section 2, we have $P(x,k^*,\beta^*) \le P(x^*,k^*,\beta^*)\ \forall\ x \in N(x^*)$, some neighborhood of x^*, because x^* is a local unconstrained maximizer of $P(x,k^*,\beta^*)$. Also, because (x^*,λ^*) satisfies complementary slackness, and since x^* is feasible,

$$P(x^*,k^*,\beta^*) = f(x^*) - \sum_{i=1}^{m} \beta^*_i[exp(k^*g_i(x^*))-1] = f(x^*) \le P(x^*,k,\beta)$$

$\forall\ k \ge 0,\ \beta \ge 0$. Thus, $\forall\ x \in N(x^*)$, $k,\beta \ge 0$,

$$P(x,k^*,\beta^*) \le P(x^*,k^*,\beta^*) \le P(x^*,k,\beta).$$

This shows that for each fixed k^* sufficiently large, the function

$$P(x,k,\beta) = f(x) - \sum_{i=1}^{m} \lambda_i/k\ [exp(kg_i(x))-1]$$

has a local saddle point at (x^*,k^*,β^*). Fix k^* sufficiently large and define a dual function, $\psi(\beta)$, as

$$\psi(\beta) = \max_{x \in N(x^*)} P(x,k^*,\beta).$$

Then we have

$$f(x^*) = \min_{\beta \ge 0} \psi(\beta) = \psi(\beta^*).$$

We note here that the dual problem is precisely the problem of determining an exact penalty function for the primal. In other words, to solve the dual, it is necessary to find a β^* such that x^* is a local solution to $\max P(x,k^*,\beta^*)$, provided k^* is sufficiently large. Fletcher [5] has obtained at least a theoretic and possibly a computational advance in penalty function techniques by presenting exact penalty functions for equality-constrained problems. However, there has to date appeared no computationally useful type of differentiable exact

penalty function for nonconcave inequality-constrained problems. The present function $P(x,k^*,\beta)$ is no exception, since there are not acceptable known procedures for determining β^* in nonconcave cases. Most known applicable procedures would involve general cutting plane techniques in the β space, such as the Dantzig-Wolfe decomposition [6] or the procedures of Nemhauser and Widhelm [7]. These procedures, at the present state of theory, could firstly be employed only to find a β^* corresponding to a global solution to (P) and secondly they would require an infinite sequence of global optimizations of nonconcave functions, which would appear to be unacceptable.

V. Example

One point in the development thus far should perhaps be further emphasized. We have shown that if x^* is a local solution to (P), and if (x^*,λ^*) satisfy the conditions D_1 thru D_4, then x^* is a local unconstrained maximizer of $P(x,k^*,\lambda^*/k^*)$ for all k^* sufficiently large. However, the theory admits the possibility of nonuniqueness. That is, if $\hat{x}$ is any other local unconstrained maximizer of $P(x,k^*,\lambda^*/k^*)$ such that $g(\hat{x}) \leq 0$ and $\lambda^*_i g_i(\hat{x}) = 0$, $i = 1,\ldots,m$ then $\hat{x}$ is a local solution to (P). We have not been able to impose weak conditions which rule out this possibility. However, note that if such an $\hat{x}$ exists then from $\lambda^*_i g_i(\hat{x}) = 0$, $i = 1,\ldots,m$, it follows that $\lambda^*_i > 0$ implies $g_i(\hat{x}) = 0$. That is, $i \in B$ implies $g_i(\hat{x}) = 0$. Also,

$$\nabla f(\hat{x}) = \sum_{i \in B} \lambda^*_i \nabla g_i(\hat{x}).$$

This means that the optimal dual variables λ^* corresponding to x^* are also optimal for $\hat{x}$. For nonlinear problems, though this is clearly possible, it intuitively seems "unlikely".

As an illustration of Theorem 1, consider the following simple quadratic example.

$$\max x^2, \quad \text{s.t.}$$

$$x - 2 \le 0$$

$$-x - 1 \le 0.$$

In this case, there are two local solutions, namely -1 and $+2$. The global solution is $x^* = 2$, with corresponding optimal dual variables $\lambda^*_1 = 4$, $\lambda^*_2 = 0$. That is, (x^*,λ^*) satisfy the Kuhn-Tucker conditions. The function $P(x,k^*,\lambda^*/k^*)$, is given by

$$P(x,k^*,\lambda^*/k^*) = x^2 - \frac{4}{k^*}(exp(k^*(x-2))-1).$$

Note that here $S = R$, $S^\perp = \{0\}$, D_4 is trivially satisfied, and Theorem 1 holds for $k^* > K = \frac{1}{2}$. It is not difficult to verify that $P(x,k^*,\lambda^*/k^*)$ has the shape illustrated by Figure 1.

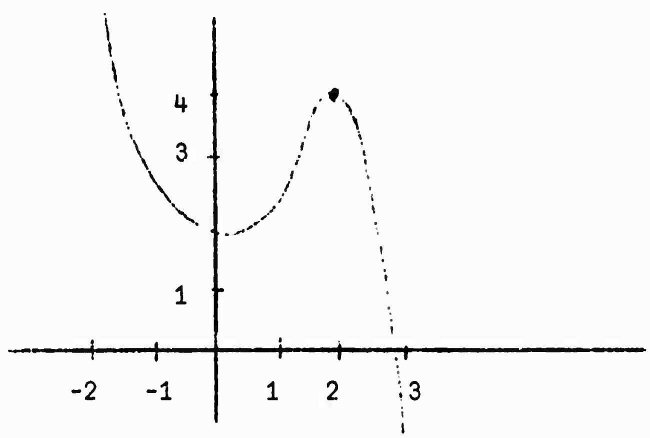

Figure 1

Although the function P is not concave in x (for any k), it is to be noted that in this case there is only one local unconstrained maximum and this occurs at $x^* = +2$. Note also that the function P has no global maximum.

VI. APPLICATION

It is generally thought to be the case that if the optimal response function is differentiable then the partial derivatives are given by the optimal dual variables λ^*_j. However, for nonconcave programs, a proof of this result, under reasonable assumptions, is not known. In this section, we offer such a proof under the following assumptions.

A_1: x^* is a unique global solution to (P).

A_2: the assumptions of Theorem 1 hold.

A_3: the following stability conditions hold (see [1])

(i) $\exists\ \delta > 0 \ni S_{\delta\bar{1}}$ is compact, where $\bar{1}$ is the m-vector of ones

(ii) $\{x \in R^n: g(x) < 0\}$ is nonempty and the closure of this set is equal to S_0, the constraint set in (P).

A_4: the optimal response is differentiable at $b = 0$ (it is not known whether this assumption is redundant to A_1, A_2 and A_3).

The optimal response function can be precisely defined as follows. Let

$$S_b = \{x \in R^n: g(x) \leq b\}, \text{ where } b \in R^m,$$

and let

$$B = \{b \in R^m: S_b \neq \emptyset\}$$

The optimal response function $f_{sup}: B \to R \cup \{+\infty\}$ is defined as

$$f_{sup}(b) = \sup f(x), \text{ s.t. } x \in S_b.$$

THEOREM 2: Under the assumptions A_1 thru A_4,

$$\left.\frac{\partial f_{sup}}{\partial b_j}\right|_{b=0} = \lambda^*_j, \quad j = 1, \ldots, m.$$

PROOF: By Theorem 1, x^* is an isolated local maximizer of $P(x,k^*,\beta^*)$, where $\beta^* = \lambda^*/k^*$ and k^* is sufficiently large. Hence, $\exists\ \varepsilon > 0$ such that

$P(x^*,k^*,\beta^*) > P(x,k^*,\beta^*)\ \forall\ x \ni ||x - x^*|| \le \varepsilon$. Let $D = \{x\epsilon R^n\colon ||x-x^*|| \le \varepsilon\}$, and consider the problem

$$(\hat{P}) \qquad \max_{x \in D} f(x)$$

$$g(x) \le 0.$$

Let $\hat{f}_{sup}(b) = sup f(x)$, s.t. $g(x) \le b$, $x\epsilon D$. We first observe that x^*, k^*, β^* satisfy the following three conditions:

(i) $P(x^*,k^*,\beta^*) > P(x,k^*,\beta^*)\ \forall\ x \in D$

(ii) $g(x^*) \le 0$

(iii) $\beta^*_i[exp(kg_i(x^*))-1] = 0, \quad i = 1,\dots,m.$ (By G_3.)

Consequently, x^*, k^*, β^* solve the extended constrained Lagrangian problem discussed in [2]. It follows from results in the latter paper that the function

$$z(b,k^*,\lambda^*) \triangleq \sum_{j=1}^{m} \lambda^*_j/k^* \, [exp(k^*b_j)-1] + \hat{f}_{sup}(0)$$

is a support to $\hat{f}_{sup}(\cdot)$ at $(0, \hat{f}_{sup}(0))$ in the sense that $\hat{f}_{sup}(b) \le z(b,k^*,\lambda^*)\ \forall\ b\epsilon B$. If we can now show that $\hat{f}_{sup}(b) = f_{sup}(b)$ for all b sufficiently near zero, it will follow that, for $j \in \{1,\dots,m\}$,

$$\frac{\partial f_{sup}(0)}{\partial b_j} = \frac{\partial \hat{f}_{sup}(0)}{\partial b_j} = \frac{\partial z}{\partial b_j}(0,\ k^*,\ \lambda^*) = \lambda^*_j$$

and the proof will be complete.

To show the required result, let $N_{1/n}(0)$ denote a family of balls of radius $1/n$ about the origin in R^m and suppose that in each of these neighborhoods there is a point b_n such that $\hat{f}_{sup}(b_n) \ne f_{sup}(b_n)$. We note that since $b_n \to 0$, the sets S_{b_n} are eventually compact and nonempty and $S_{b_n} \to S_0$ in the Hausdorff metric. It was shown in [1] that these results are implied by the stability conditions A_3. By the eventual compactness of the sets S_{b_n}, for all n sufficiently large there is an $x_n \epsilon S_{b_n}$ such that $f(x_n) = f_{sup}(b_n)$. Now, since $\hat{f}_{sup}(b_n) \ne f_{sup}(b_n)$, it follows that none of the elements x_n are

in D. Hence a subsequence of x_n converges to some $\tilde{x} \epsilon S_0$ such that $||\tilde{x} - x^*|| > \varepsilon$. But since $f_{sup}(b_n) \to f_{sup}(0)$, we have

$$f(\tilde{x}) = \lim f(x_{n_j}) = \lim f_{sup}(b_{n_j}) = f_{sup}(0) = f(x^*),$$

which contradicts the uniqueness of $f(x^*)$. Thus we have $\hat{f}_{sup}(b) = f_{sup}(b)$ for all b in some neighborhood of zero. □

References

[1] J. P. Evans and F. J. Gould, "Stability in nonlinear programming," *Opns. Res.*, *18* (1970), pp. 107-118.

[2] F. J. Gould and Jon W. Tolle, "A necessary and sufficient constraint qualification for constrained optimization," *SIAM J. Appl. Math.*, to appear.

[3] F. J. Gould, "Extensions of Lagrange multipliers in nonlinear programming," *SIAM J. Appl. Math.*, *17* (1969), pp. 1280-1297.

[4] A. V. Fiacco and G. P. McCormick, *Non-linear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley and Sons, New York, 1968.

[5] R. Fletcher, *A Class of Methods for Non-linear Programming with Termination and Convergence Properties*, *Integer and Nonlinear Programming*, J. Abadie, ed., North-Holland Publishing Company, Amsterdam, 1970.

[6] G. B. Dantzig, *Linear Programming and Extensions*, Princeton University Press, New Jersey, 1963.

[7] G. L. Nemhauser and W. B. Widhelm, "A modified linear program for columnar methods in mathematical programming," *Opns. Res.*, to appear.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

Department of Statistics
University of North Carolina
Chapel Hill, N.C. 27514

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

3. REPORT TITLE

A New Result on Interpreting Lagrange Multipliers as Dual Variables

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Research Report

5. AUTHOR(S) (First name, middle initial, last name)

F. J. Gould and Stephen Howe

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
January, 1971	-15-	-07-

8a. CONTRACT OR GRANT NO.

N00014-67-A-0321-003 (NR047-095)

b. PROJECT NO.

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)

Institute of Statistics Mimeo Series
Number 738

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

The distribution of this report is unlimited

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

Operations Research Program (Code 434)
Office of Naval Research
Washington, D.C. 20360

13. ABSTRACT

It is well known that in linear programming the knowledge of optimal dual variables can be used to simplify the finding of an optimal solution to the primal problem. In this paper, the analogous notion is considered for nonlinear and particularly nonconcave programming problems. A new theoretic result on Lagrange multipliers is presented which enables one to make a statement for nonlinear problems similar to the above assertion for linear programs.

Unclassified
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Dual Variables						
Nonconcave Programs						
Lagrange Multipliers						
Optimization Theory						

Unclassified
Security Classification